SPAWAR
Systems Center
San Diego

TECHNICAL REPORT 1917
July 2004

GeoPlot Declutter Interface Usability Evaluation

Bela Feher

SSC San Diego

TECHNICAL REPORT 1917
July 2004

GeoPlot Declutter Interface Usability Evaluation

Bela Feher

SPAWAR
Systems Center
San Diego

SSC San Diego
San Diego, CA 92152-5001

SSC SAN DIEGO
San Diego, California 92152-5001

T. V. Flynn, CAPT, USN
Commanding Officer

R. F. Smith
Executive Director

ADMINISTRATIVE INFORMATION

This report was prepared for the GeoPlot Declutter Project (244210) of the Simulation and Human Systems Technology Division (244) of the Command and Control Department (240) of Space and Naval Warfare Systems Center, San Diego (SSC San Diego). Funding was provided by the Office of Naval Research (ONR 342), Cognitive and Neural Science Division, under program element 0602233N. The ONR program officer was Gerald Malecki.

Released by
J. G. Morrison, Principal Investigator
Collaborative Technologies Branch

Under authority of
J. L. Martin, Head
Simulation and Human
Systems Technology Division

SB

ACKNOWLEDGMENTS

This report was prepared as part of SSC San Diego's ongoing Command 21 research program, sponsored by the Office of Naval Research, Cognitive and Neural Science Program, with Gerald Malecki as Program Officer. Dr. Jeffrey G. Morrison, SSC San Diego, is the principal investigator on the Command 21 project. The GeoPlot Declutter Project investigates methods of intelligent attention management for reducing clutter and helping users focus on critical information. These methods require basic research in three related fields:

- Algorithms for evaluating and identifying air and surface threats.
- General methods for directing users' attention toward high threats without reducing global situation awareness.
- The design of human–automation interfaces that help users supervise and interact with realistically imperfect automated systems with sophistication and efficiency.

The ultimate goal is to produce interface design principles and construct an integrative model of human–automation interface design that promotes sophisticated use of automation in producing and maintaining situation awareness.

The usability evaluation reported here is part of a larger experiment conducted as the culmination of a great deal of development, testing, and evaluation over several years. The full experiment is reported in detail in St. John, Manes, Smallman, Feher, & Morrison, 2004, Intelligent Decluttering of Tactical Displays to Improve Operator Performance. The author would like to thank the following individuals who contributed directly to the design and conduct of the current study:

- Dr. Mark St. John, Mr. Dan Manes, Dr. Harvey Smallman, Pacific Science and Engineering Group, Inc.
- Dr. Sandra Marshall, San Diego State University
- Dr. Jeffrey G. Morrison, SPAWAR Systems Center, San Diego

The author would like to thank the following individuals whose groundbreaking earlier efforts and contributions provided the foundations for the current work:

- Mr. Ronald Moore and Mr. Gene Averett

EXECUTIVE SUMMARY

This study evaluates the usability of an advanced human–system interface that uses an automated algorithm to reduce user workload by automatically decluttering low-threat tracks from a tactical air warfare display. The algorithm assists users by pre-classifying the threat level of each contact and desaturating (dimming) the symbols of less threatening contacts while keeping the symbols of those evaluated as significant threats at full brightness. The usability evaluation was conducted in association with an experiment designed to determine the value of GeoPlot declutter in an operational setting (St. John et al., 2004). The experiment had two purposes: (1) to compare two decluttered interfaces with a standard, non-decluttered interface for performing an air defense task in a simulated operational setting, and (2) to compare medium and high threshold cutoffs on the declutter algorithm. (The high threshold effectively results in decluttering low-threat tracks and "borderline" threat tracks while medium threshold only declutters the low threats.)

This study is the culmination of several previous studies along three major lines of research: (1) threat classification algorithm development based on the track parameters used by experts in classifying threats, (2) attention management studies directed at packaging multi-parameter information in symbology and the performance benefits of varying the saturation of symbols in cluttered GeoPlots, and (3) interface design studies based on user-centered design principles to improve decision performance of warfare commanders in operational settings.

Participants with air defense warfare expertise were asked to use and evaluate three GeoPlot interfaces in performing an air defense task as an Air Warfare Commander whose responsibility was to monitor all air traffic in a littoral situation and take protective actions dictated by the Rules of Engagement to defend the high-value platform. After performing the task with each of the three interfaces, they were asked to provide usability evaluations of the three interfaces and provide any criticisms or suggestions for improvement. These evaluations were collected in writing and verbally.

Participants' post-task task difficulty ratings clearly indicated that the declutter interfaces were helpful in reducing task difficulty compared to the more traditional no declutter interface. However, their reported confidence that they performed well was unaffected by the interfaces.

Participants' subjective estimates of performance after experience in the scenarios were moderately high on three different measures: threat detection, change in threat status, and ability to maintain situation awareness. In all cases, their subjective estimates of performance were higher for the declutter interfaces than for the no declutter interface. These estimates were supported by many verbal references to the beneficial impact of the declutter interfaces on workload.

When participants were asked to compare the declutter interfaces to the no declutter interface on six dimensions of usability, they gave the medium threshold declutter interface a greater advantage over no declutter than they gave the high threshold declutter interface. This response is consistent with their ratings of usefulness of the interfaces, where both declutter interfaces were much more highly rated than the no declutter interface. It is also consistent with their preference rankings of the interfaces, where medium threshold declutter was strongly preferred to high threshold declutter, which was preferred to no declutter.

For automation trust, participants were clearly aware of the issue from a practical standpoint. They made many comments about agreement or disagreement with the algorithm, having to double-check its results, and its value in helping or hindering their attention to important tracks. We used two measures to tap their inclinations to trust or distrust the classification algorithm: (1) accuracy of highlighting, and (2) the need to double-check track classifications. The participants judged the

accuracy of the algorithm in highlighting significant threats as quite high, possibly providing a necessary basis for trust. The second measure tapped the participants' need to double-check the results of the algorithm by hooking tracks and making their own assessments. According to their reports on the need to double-check the automation, participants did not yield their trust completely to the automation, but reported that regardless of the declutter threshold or the automated classification, they felt it necessary to check "some" of the tracks, as opposed to "few" or "most." These results, along with participants' comments about their training- and experience-based behavior patterns and their declutter usage strategies suggest that operationally experienced users appreciate the workload and prioritization benefits of automation while realistically tempering their reliance on it.

Participants claimed various creative strategies for using declutter capabilities. These strategies are probably subject to further refinement with extended experience, and they may take a different form if additional declutter interface features that they requested are added. Many additional features were suggested, and both these suggestions and many verbal comments indicate that the declutter interfaces are worthy of further development.

This study, which simulated an air defense task in a dense littoral operational environment, produced the following conclusions:

- Declutter interfaces were judged superior in overall usefulness compared to no declutter.
- Automation trust is realistic in this community, while users are highly receptive to ways to prioritize their work and reduce their workload.
- Declutter interfaces rated higher usability ratings than the no declutter interface on all measured dimensions.
- This evaluation of declutter interfaces compared with a no declutter interface established a clear mandate to build on the tested declutter concepts.
- Further development should incorporate additional features and tools suggested by these experienced air warfare personnel.

CONTENTS

Figures

INTRODUCTION

PURPOSE

Theaters of operation are busy environments, and displays of tactical situations can quickly become congested and cluttered with track symbols. Severe clutter can have many deleterious effects. Important information may be obscured or masked, irrelevant information may receive undue attention, and cognitive workload may escalate unnecessarily—all leading to difficult threat assessment and less than optimal situation awareness. This project developed an experimental declutter tool for single ship air and surface warfare. The tool uses a neural network-based algorithm to model human threat evaluation and decision-making. Information about individual tracks is fed into the model, and the model produces a "level-of-interest" score for that track. The algorithm assists users by pre-classifying the threat level of each contact and desaturating (dimming) the symbols of less threatening contacts while keeping the symbols of those evaluated as significant threats at full brightness.

This study evaluates the usability of an advanced human–system interface that uses an automated algorithm to reduce user workload by automatically decluttering low-threat tracks. The usability evaluation was conducted in association with an experiment conducted for two purposes: (1) to compare two decluttered interfaces with a standard, non-decluttered interface for performing an air defense task, and (2) to compare medium and high threshold cutoffs on the declutter algorithm. The behavioral data from the experiment are reported in St. John et al., 2004.

BACKGROUND

This study is the culmination of several previous studies along three major lines of research: (1) threat classification algorithm development based on the track parameters used by experts in classifying threats, (2) attention management studies directed at packaging multi-parameter information in symbology and the performance benefits of varying the saturation of symbols in cluttered GeoPlots, and (3) interface design studies based on user-centered design principles to improve decision performance of warfare commanders in operational settings.

Threat Classification Parameters of Experts

Marshall, Christensen, and McAllister (1996), Liebhaber (2001), and Liebhaber, Kobus, and Feher (2002) determined the threat classification parameters used by experienced air defense personnel. These parameters were used to construct an algorithm to automatically "pre-classify" unknown tracks (those that are somewhat ambiguous and threatening) for the user. The algorithm does not consider "hostiles," which would receive scores greater than 10. Friendly contacts are assigned a score of 0. The algorithm produces a threat score of 0 to 10 for each track, each second. According to the Rules of Engagement (ROE), scores of 8 or above are significant threats.

User Interface Symbology and Desaturation Studies

St. John, Feher, and Morrison (2002) and Smallman et al. (2001a; 2001b) studied symbology as means of conveying threat-related information about each track and desaturation as a means of managing the user's attention in a cluttered display. The interface design used in this experiment consisted of MIL–STD–2525 symbols for air contacts, using yellow for unknown tracks and blue for friendly tracks. All tracks were continuously displayed, but the symbols of low-threat tracks were desaturated by decreasing their brightness to 30%. These earlier studies established that workload and performance were profoundly affected by desaturation.

Tactical Decision-Making Under Stress (TADMUS) Studies

The air defense scenarios in the GeoPlot declutter study are modelled after Tactical Decision-Making Under Stress (TADMUS) scenarios (Hutchins, 1996; Hutchins, Morrison, and Kelly, 1996; Kelly, Hutchins, and Morrison, 1996) that posed a Battle Group (BG) air defense situation in a littoral setting that includes four air corridors and mixed air traffic, primarily categorized as unknown. These studies led to a new interface design for an air defense decision support system that significantly improved performance of Air Warfare Commanders. Each minute of the 15-minute GeoPlot declutter scenarios presented a progressive array of approximately 50 contacts that were evaluated by the algorithm as having scores of 0 to 5.99 (N = 31), 6 to 7.99 (N = 12), or 8 to 10 (N = 7).

METHOD

Participants with air warfare expertise were recruited from training command staffs and a carrier air wing. The criterion for participant recruitment was knowledge of air defense operations and familiarity with the ROE.

Participants were provided a Voluntary Consent Form for their signature, a Privacy Act Statement, and a Demographic Information Form that asked relevant information on their rank, training, experience, and qualifications.

Participants were shown the three GeoPlot interfaces that they would use in the study, and the major features and differences were explained. They were then given a 3-minute practice session with the non-decluttered interface. During this session, they were encouraged to try out the interface features and ask any questions they had about the interface or scenarios.

Following the brief training on the task using the non-decluttered interface, participants were asked to perform an air defense task during three randomly ordered scenarios. These scenarios presented a typical littoral air defense picture where a high-value platform was operating in the midst of multiple airlanes and relatively dense air traffic representing friendly and unknown contacts. As Air Warfare Commander, the participant's job was to monitor all air traffic and take actions dictated by the ROE. The ROE required that all high-threat tracks be acted on as follows:

- At 75 miles, notify the BG Commander
- At 50 miles, query the contact for identification
- At 25 miles, warn the contact to change course

Optional actions allowed were illuminating the contact and requests to engage the contact. All notifications, queries, and warnings were recorded and timed with respect to delay after crossing the distance threshold set for each action by the ROE. Participants were told that they were the experts on threat classification, even when the interface provided automated assistance, so their judgment was the ultimate determinant of threat classification. They were instructed that based on their judgments of threat, they were to act on all high-threat tracks according to the ROE requirements.

Confidence and difficulty judgments were collected as follows: (1) immediately after introduction of the task, and (2) after introduction of the three interfaces and the practice trial. After each of the three 15-minute scenarios, during which participants gained experience with each interface (and data were collected on their performance), participants filled out the NASA Task Load Index (TLX) scale of subjective workload. After the third performance trial, they provided usability evaluations of the three interfaces and any criticisms or suggestions for improvement. These evaluations were collected in writing and verbally.

SUBJECTS

The participants were 27 naval personnel, 26 male and 1 female. Ages ranged from 24 to 54 years, with a mean of 35 years. Eight participants were chiefs or senior chiefs from the Aegis Training and Readiness Center Detachment, San Diego (E-7 to E-8); three were senior officers from the Tactical Training Group, Pacific (O-5 to O-6); and 16 were junior officers (O-2 to O-4) from the Airborne Early Warning Wing, Pacific. The participants had from 3 to 30 years of service in the U.S. Navy, with an average of 13 years. Air warfare expertise and experience was rated on a three-point scale for each participant by a subject matter expert. Fourteen of the participants were given a very high rating, two were given a high rating, and eleven were given a moderate rating.

PROCEDURE

The study components were administered to each participant in the following order:

- Voluntary Consent Form and Privacy Act Statement
- Collection of demographic and training/experience data
- Orientation—general task description using non-decluttered GeoPlot display, with hands-on demonstration of tools and actions
- Specific Tasking—verbal task description of Air Warfare Commander responsibilities, described using non-decluttered interface on a static GeoPlot display, with hands-on trial of features
- Initial user interface evaluation questions on expected task difficulty and confidence (Pre-Measure 1)
- Introduction to decluttered interface versions (static views of high threshold and medium threshold interfaces).
- 3-minute practice session using dynamic, non-decluttered interface
- User interface evaluation questions on expected difficulty and confidence using each interface (Pre-Measure 2)
- Trial 1 (one of three interface conditions randomly ordered)
- Workload questionnaire (TLX)
- Trial 2 (second of three interface conditions)
- Workload questionnaire (TLX)
- Trial 3 (third of three interface conditions
- Workload questionnaire (TLX)
- User interface questions evaluating task confidence and difficulty, ratings and comparisons of interfaces, and their criticisms and suggestions for interface design (Post-Measure)

For three participants at Tactical Training Group Pacific (TTGP), a modified study design was used to incorporate eye-tracking data collection. Eye-tracking was expected to add the following: (1) a physiological measure of cognitive workload, and (2) a capability to analyze the scanning patterns during processing of information tracks. In this part of the study, it was decided to eliminate the TLX. This decision was based on two factors: (1) the TLX overlapped with the eye-tracking workload measure, and (2) participants would not be able to fill out the TLX questionnaire with the eye-tracking headgear installed, which would require re-installation and calibration that would substantially lengthen the time required of participants. The modified procedure consisted of the following:

- Voluntary Consent Form and Privacy Act Statement
- Collection of demographic and training/experience data
- Orientation—general task description using non-decluttered GeoPlot display, with hands-on demonstration of tools and actions

- Specific Tasking—verbal task description of Air Warfare Commander responsibilities, described using non-decluttered interface on a static GeoPlot display, with hands-on trial of features
- Initial user interface questions on expected task difficulty and confidence (Pre-Measure 1)
- Introduction to decluttered interface versions (static views of high threshold and medium threshold interfaces).
- 3-minute practice session using dynamic, non-decluttered interface
- User interface questions on expected difficulty and confidence using each interface

(Pre-Measure 2)

- Trials 1, 2 , and 3 (three interface conditions randomly ordered)
- User interface questions evaluating task confidence and difficulty, ratings and comparisons of interfaces, and their criticisms and suggestions for interface design (Post-Measure)

RESULTS

ESTIMATED TASK DIFFICULTY AND CONFIDENCE BY INTERFACE

Task Difficulty (Q1, Q3)

As the participants were introduced to the scenario and their task, several estimates of their perceptions of task difficulty and their confidence in performing well were obtained. The first measure of task difficulty was obtained immediately after a description of the general scenario while viewing a static GeoPlot screen depicting a fairly dense littoral environment. Their task was summarized as "main-taining situation awareness and responding to significant air threats in a timely manner" (Q1). Participants' mean rating of task difficulty was 2.26 on a five-point scale, where 1 was Low Difficulty and 3 was Moderate Difficulty.

A second measure of task difficulty was obtained immediately after the introduction of the three GeoPlot interfaces that they would use to perform their task and hands-on experience with the no declutter interface. The introduction of the interfaces and practice resulted in a recalibration of participants' estimates of task difficulty, with no declutter estimated to be moderately difficult (mean = 2.90; $t_{Inital,\,N} < 0.05$), medium threshold declutter judged somewhat less difficult (mean = 2.24; $t_{N,M} < 0.05$), and high threshold declutter rated least difficult (mean = 1.94; $t_{N,H} < .05$; $t_{M,H} < 0.05$)). Their expectations following introduction to the declutter interfaces were clearly that they expected decluttering to reduce the task difficulty.

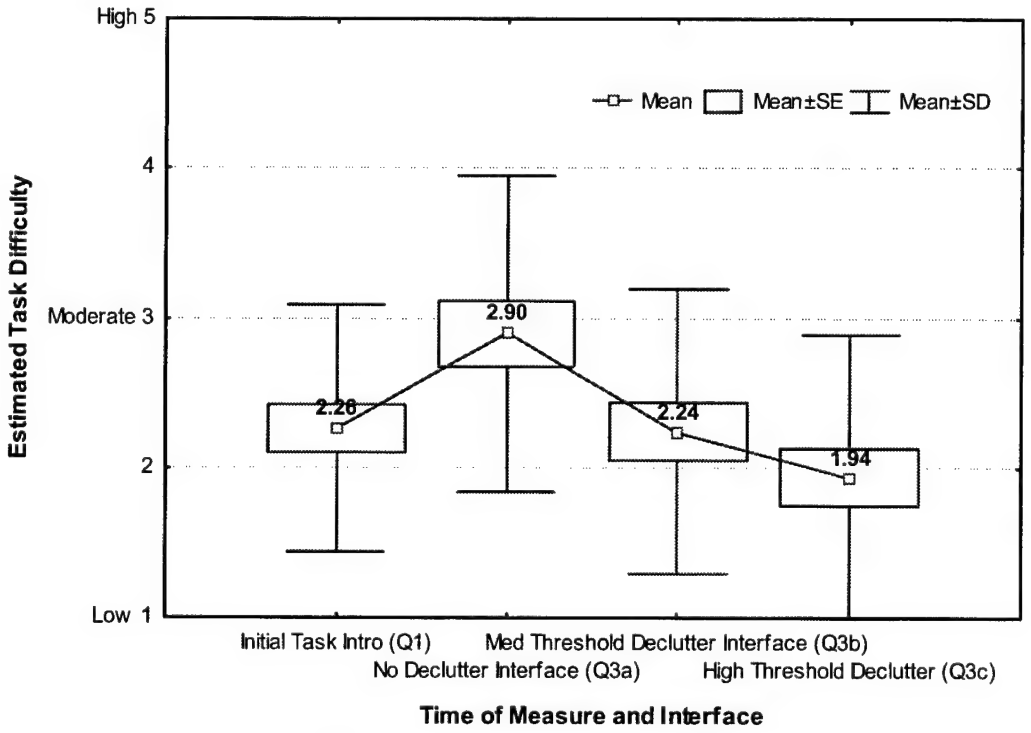

Figure 1. Pre-task measures of task difficulty.

Confidence (Q2, Q4)

When the measures of task difficulty were taken, participants were asked their confidence in performing the task well. Their initial confidence rating on exposure to the littoral scenario was moderately high (mean = 3.94 on a five-point scale). Again, they appeared to recalibrate their confidence estimates upon exposure to the static interfaces and the practice experience. They adjusted their confidence somewhat downward for the no declutter interface (mean = 3.65; $t_{Inital, N} < 0.05$). Their confidence when using the medium threshold declutter or the high threshold declutter interfaces was significantly higher than with the no declutter interface ($mean_M = 4.01$ and $mean_H = 4.16$; $t_{N, M} < 0.05$ and $t_{N, H} < 0.05$). As with task difficulty, participants' expectations were that the declutter interfaces would help them improve performance relative to the traditional no declutter interface.

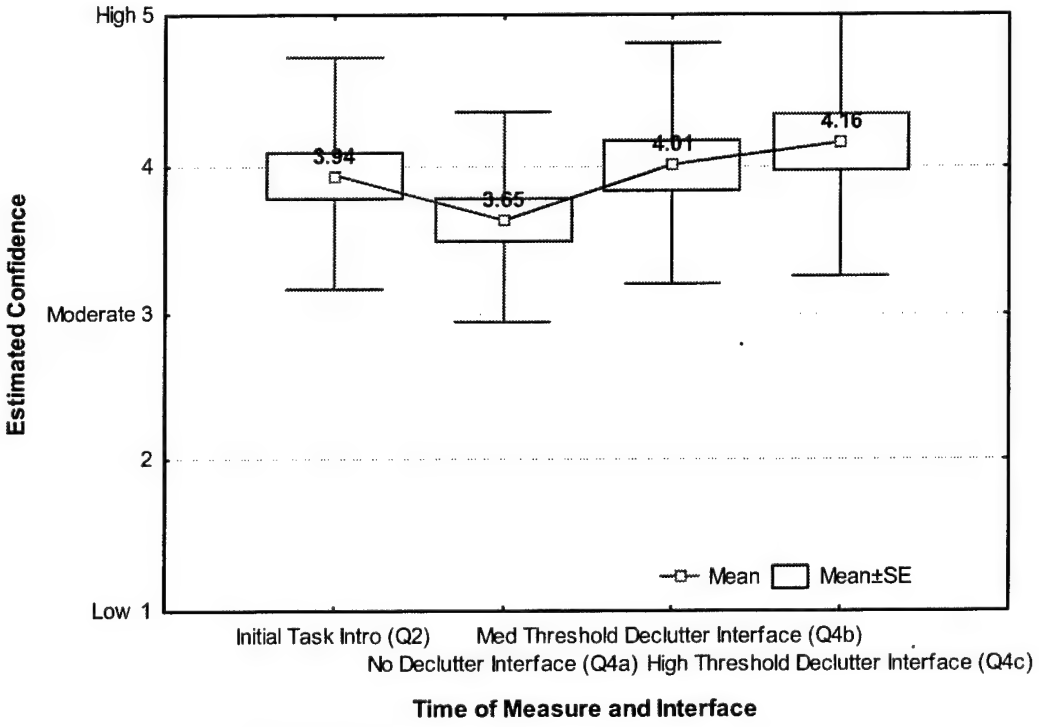

Figure 2. Pre-task measures of confidence.

Post-Task Difficulty Ratings (Q5)

After the three 15-minute scenarios, during which they used each interface to perform the air defense task, participants again rated task difficulty with each interface. They reported the task as moderately difficult when using the no declutter interface (mean = 2.88), but low to moderate when using either of the two declutter interfaces ($mean_M = 2.04$ and $mean_H = 2.10$; $t_{N,M} < 0.05$ and $t_{N,H} < 0.05$). Thus, participants reported that the declutter interfaces were helpful in reducing task difficulty compared to the traditional no declutter interface.

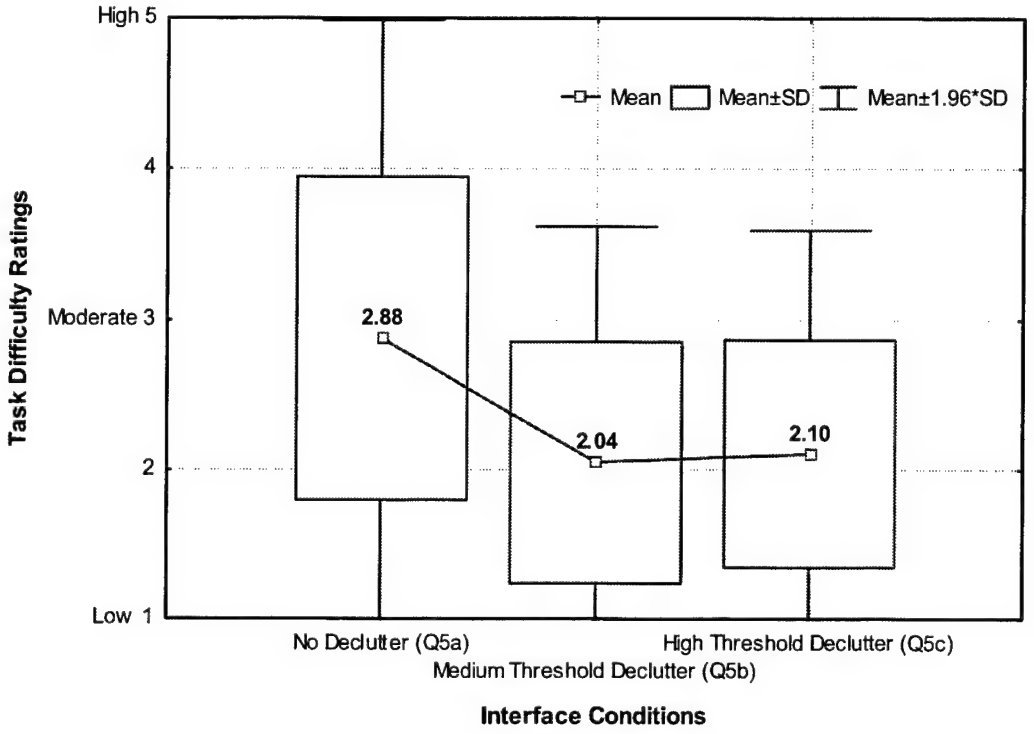

Figure 3. Post-task task difficulty ratings by interface.

Post-Task Confidence Ratings (Q6)

Participants also provided their confidence ratings after completion of the three task scenarios in which they used the different interfaces. They reported moderately high confidence in their performance using all three interfaces, with no significant differences among their ratings ($mean_N = 3.92$; $mean_M = 4.19$, and $mean_H = 4.11$). These confidence ratings apparently reflect their perceptions of having performed well during the scenarios.

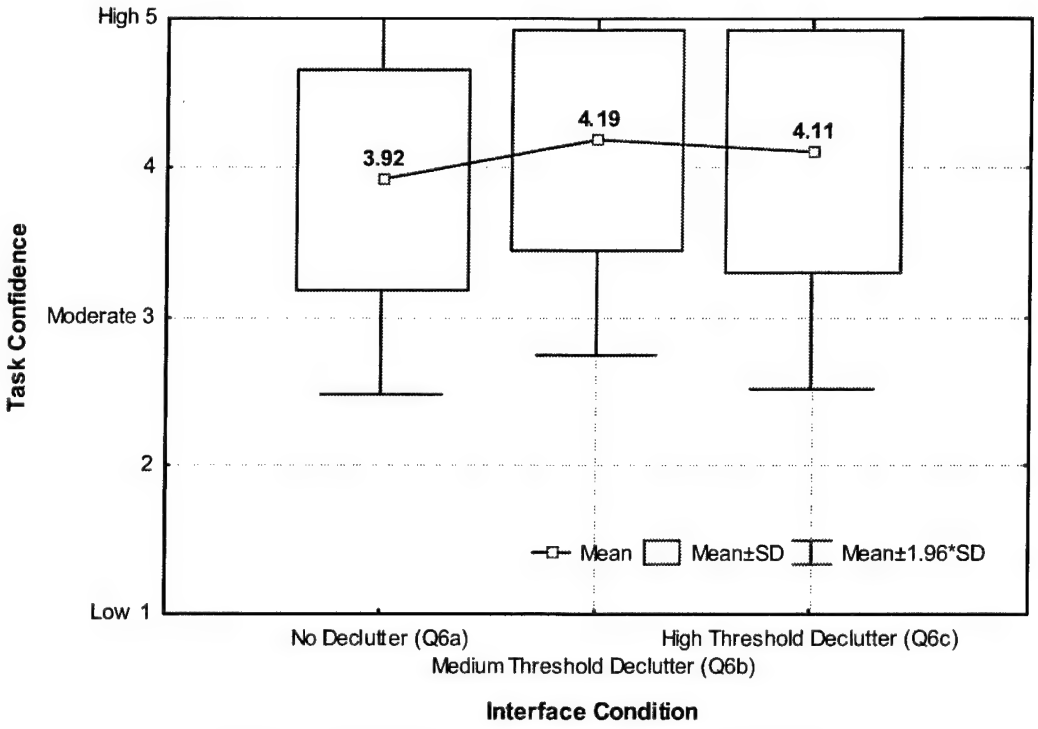

Figure 4. Post-task confidence ratings by interface.

EXPERIMENT CONDITIONS

As a validation of the experimental conditions, it was important to obtain participants' evaluation of the scenario and air warfare task that they were asked to perform. These task evaluations were common to all interface conditions.

Scenario Realism (Q16a)

The scenario was constructed as a dense littoral air defense situation where hostile and friendly nations were in close proximity, with ComAir lanes nearby and heavy air traffic in the vicinity (Figures 5a and 5b). Participants' mean rating of the realism of the scenario was slightly above moderate (3.49 on a five-point scale).

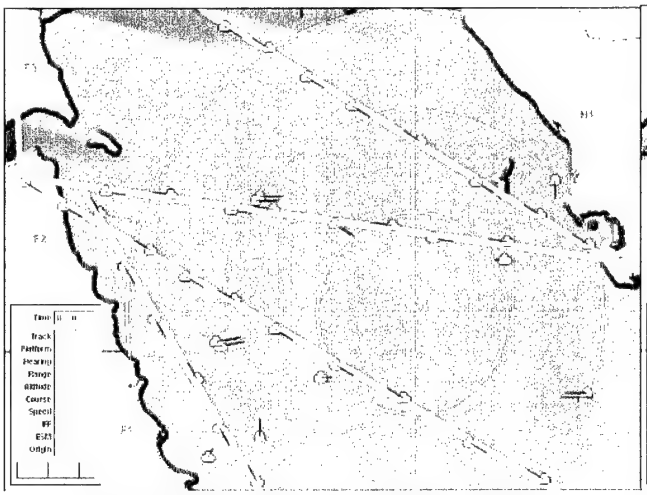

Figure 5a. No declutter GeoPlot display.

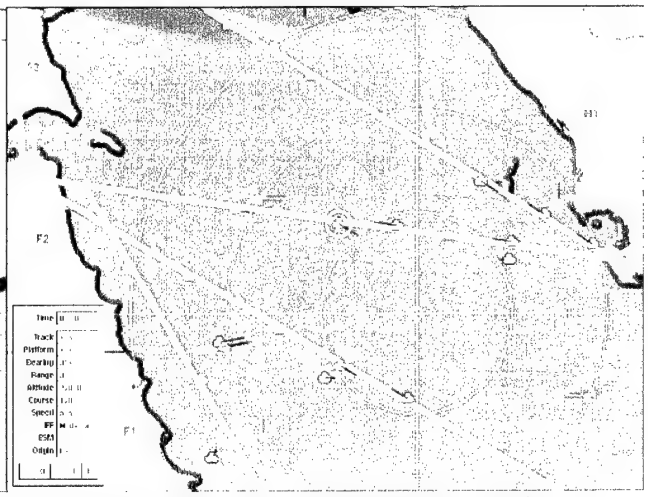

Figure 5b. Decluttered GeoPlot display.

Realism of Task Requirements (Q16b)

The second validation measure obtained from participants was their judgment of the realism of the task requirements. Participants rated the task requirements slightly above moderately realistic (mean = 3.62).

When participants made verbal comments about the scenario or task, they generally supported them as plausible and effective, although one person said that real operations were likely to involve more ships to be protected (than just ownship) and others indicated that there are several, somewhat redundant players who deal with sectors or otherwise coordinate a complex network of air defense roles in the operational setting. Although the experimental scenario simplified the situation to a single actor, they generally felt it was a plausible situation and a task that adequately reflected actual operations.

Although participants' ratings according to their experience levels did not differ significantly, participants who were classified as more experienced in air defense warfare assigned higher realism ratings to the scenarios than the less experienced participants ($mean_{Very\ High + High}$ participants = 3.73, $mean_{Moderate}$ participants = 3.14, p = 0.02). More experienced participants also tended to rate the experimental task higher in realism than less experienced participants ($mean_{VH+H}$ = 3.86; $mean_{M}$ = 3.27; p = 0.05). (Appendix B shows the figures.)

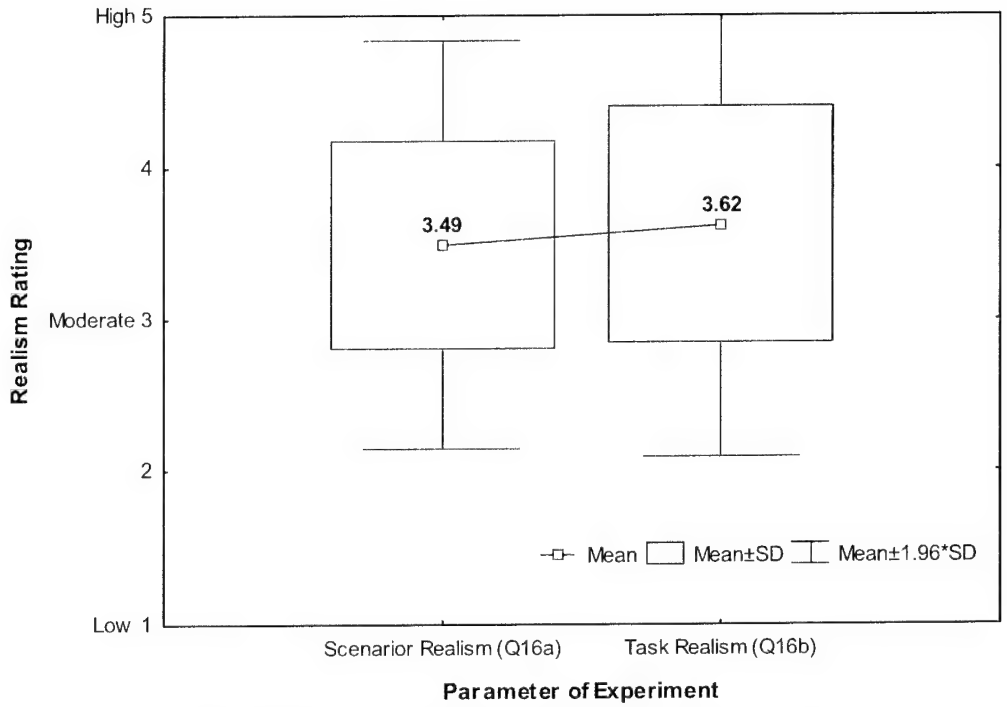

Figure 6. Validation of experiment scenario and task.

SUBJECTIVE PERFORMANCE

Threat Detection by Interface (Q7)

After using each of the three interfaces, participants rated their "ability to detect threats while using each of the interfaces." They rated their performance moderately high ($mean_N = 3.88$, $mean_M = 4.38$, and $mean_H = 4.21$). However, they rated their threat detection performance significantly lower when using the no declutter interface than when using the medium threshold declutter interface ($t_{N, M} < 0.05$). Other interface differences were not significant. This perceived difference in performance is consistent with their reported preference for the medium threshold declutter interface over the high threshold declutter and no declutter interfaces (Q14 and Q15).

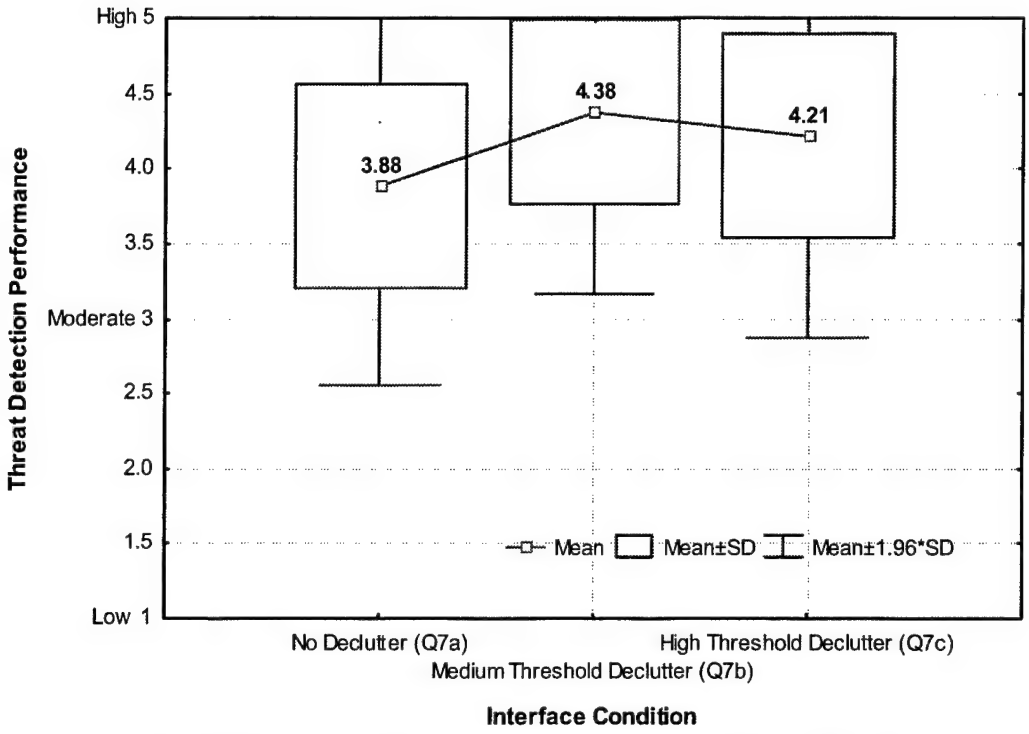

Figure 7. Subjective threat detection performance by interface.

Detection of Changes in Threat Status by Interface (Q8)

One expected benefit of the declutter interfaces is that highlighted threat tracks are distinguished from non-threat tracks (as determined by the algorithm) and draw attention to themselves by their brightness. However, the value of this simple feature to detect changes in threat status remains an empirical question. For this reason, participants were asked to report their perceptions of the effect of highlighting on their "ability to detect changes in threat status using each of the interfaces" (Q8). Without highlighting (no declutter interface (Q8a)), participants reported that they felt they could achieve slightly better than a moderate level of detecting changes in threat status (mean = 3.45). Their judgments of their performance using the declutter interfaces were higher, especially using the medium threshold declutter interface ($mean_M = 4.01$, and $mean_H = 3.89$; $t_{N,M} < 0.05$, $t_{N,H}$ = n.s.). Again, participants felt that declutter, especially using a medium threshold, could improve their detection of changes in threat status.

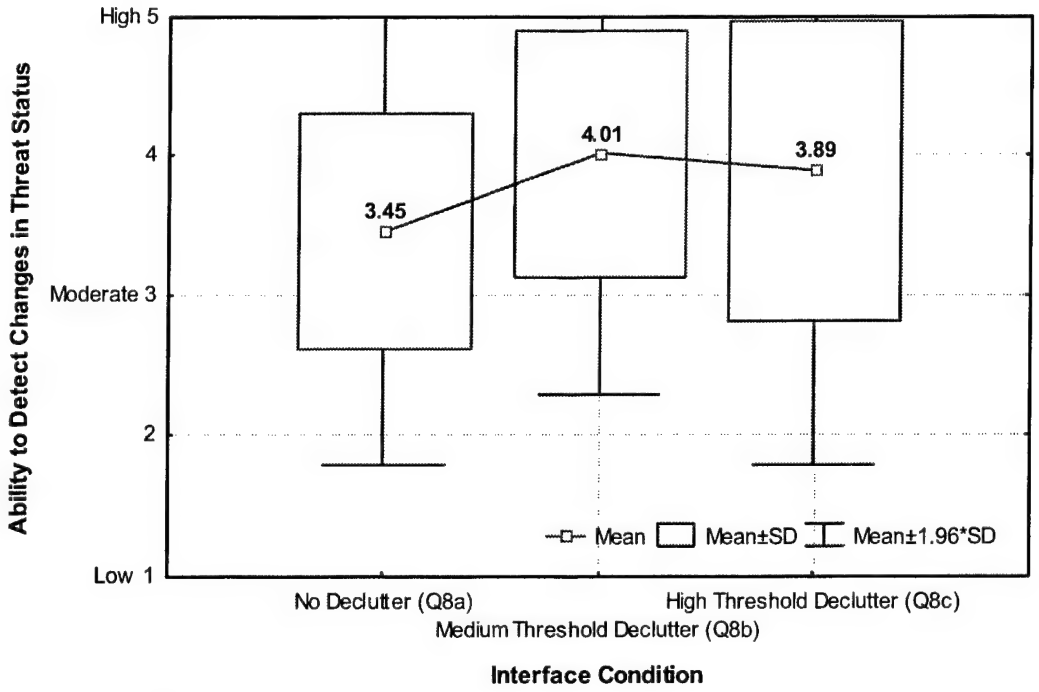

Figure 8. Subjective ratings of change detection performance by interface.

Maintaining Situation Awareness by Interface (Q9)

Another hypothesized effect of the declutter interfaces is a positive impact on the users' ability to maintain situation awareness. After the use of each interface, this effect was tapped in an item that asked them to "rate your ability to maintain overall situation awareness while using each of the interfaces." Participants reported that without decluttering (using the no declutter interface) they could maintain slightly better than moderate situation awareness (mean = 3.53). With the declutter interfaces, they reported a perceived improvement in their ability to maintain situational awareness with each, but they did not perceive any distinction in ability to maintain situational awareness between the two declutter interfaces ($mean_M = 4.20$, $mean_H = 4.19$, $t_{N,M} < 0.05$, $t_{N,H} < 0.05$, $t_{M,H}$ = n.s.). In essence, participants perceived highlighted threat tracks in the declutter interfaces as having a positive effect on users' ability to maintain situational awareness.

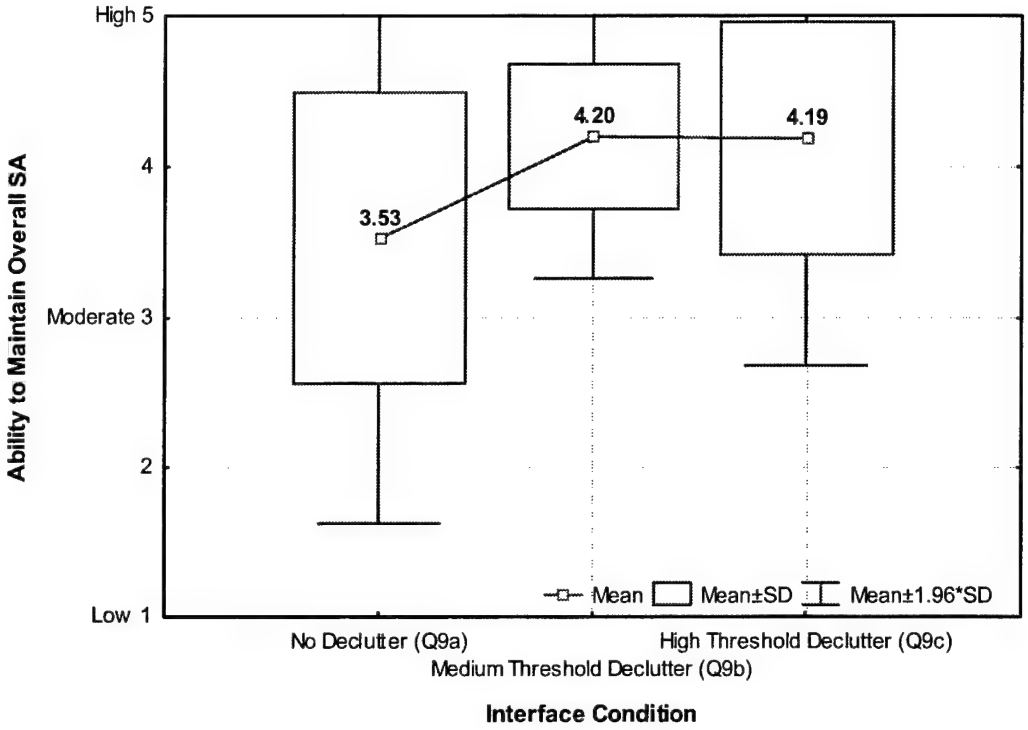

Figure 9. Ability to maintain situation awareness by interface.

TRUST IN AUTOMATION BY DECLUTTER INTERFACE THRESHOLD

Participants were told that the declutter interfaces used an imperfect algorithm for providing a "first cut" at threat classification. According to the protocol, the ROE defined "significant threats " as a "track that has the 'potential' to threaten ownship and has menacing kinematics—in other words, a track that would be rated as an 8 on a threat scale from 1 to 10." Participants were repeatedly cautioned that they were the ultimate authorities in determining what tracks were threats, regardless of the determination of the threat algorithm. As Air Warfare Commanders, participants were tasked with responding appropriately according to the ROE to each track they determined as a threat. Thus, each participant needed to decide how much trust to put in the algorithm's classifications. This decision could result in different usage patterns (strategies) and different effects on their workload.

If it were perfect, the high threshold declutter algorithm would highlight all tracks that the ROE required the Air Warfare Commander to act on, i.e., tracks with a threat rating of 8 to 10. On the one hand, when using the high threshold declutter interface, participants who had high trust in the automation could allocate their initial scanning effort to the highlighted tracks, afterward turning to the decluttered (dimmed) tracks to make sure that they were not threats. By enabling them to prioritize the tracks to which they allocated their attention, these participants might be able to reduce their workload while improving their reaction times in dealing with threats.

On the other hand, participants who did not trust the automation while using the high threshold interface needed to allocate their time equally among all tracks to make their own determinations of threat classification. They would probably not experience any reduction in workload or any improvement in reaction times in dealing with threats.

The user experience was quite different for participants using the medium threshold declutter interface. This interface highlighted tracks with threat ratings of 6 to 10, allowing for possible errors at the threat margin. This threshold still allowed the participants to focus their attention on higher priority tracks, but ruled out the tracks least likely to qualify as threats. The net effect could be to reduce workload and gain response improvements, possibly in lesser degree than the high threshold declutter interface due to the retention of more highlighted tracks. Presumably, participants who did not want to place a high degree of trust in the automation could gain performance advantage using the medium threshold declutter interface while remaining comfortable about using their own expertise to determine threat classifications.

Besides performance measures, participants were asked two sets of questions about their perceptions of the high and medium threshold interfaces: (1) the perceived accuracy of the automation in "highlighting all of the tracks you rated to be significant threats," and (2) the "proportion of tracks" they felt the need to "double-check for accuracy of classification."

Highlighting Accuracy (Q10)

Participants judged the accuracy of the algorithm in identifying threats as quite high (mean = 4.08 on a five-point scale for medium threshold and 3.98 for high threshold). These judgments were based on their experience with the two interfaces in two separate scenarios. Because the two interfaces differed only in the threshold setting for determining the tracks to be decluttered (dimmed), it is appropriate that the participants rated their accuracy as equivalent. The fact that their ratings were near the high end of the scale indicates that the participants were in substantial agreement with the threat classifications made by the algorithm.

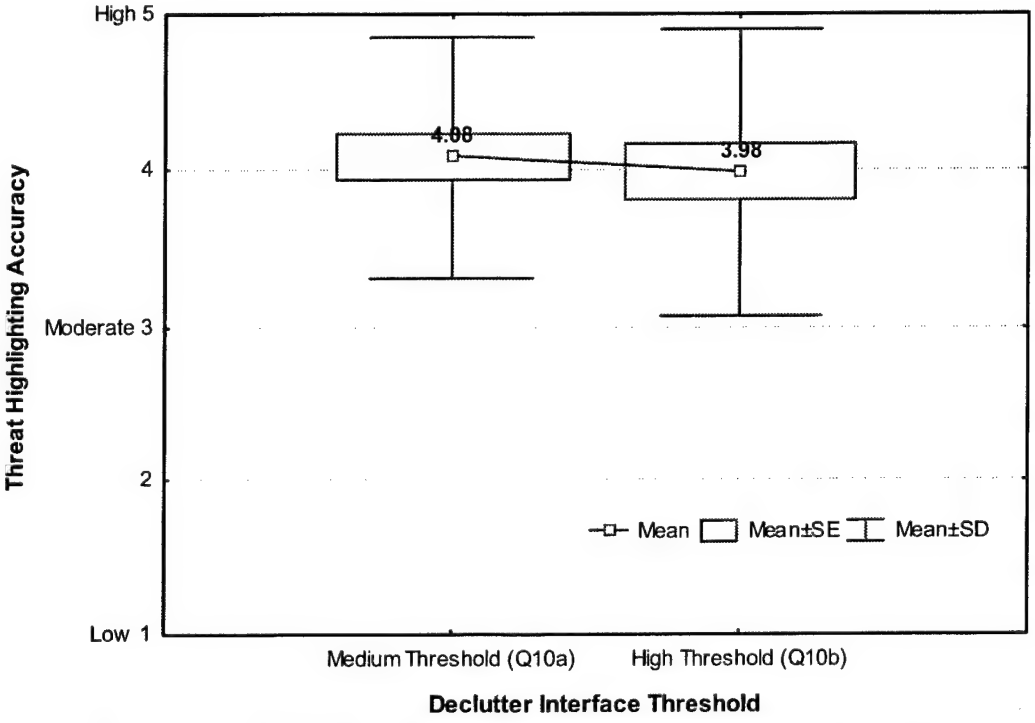

Figure 10. Highlighting accuracy of automation by declutter threshold.

Need to Double-Check by Threat Level and Threshold (Trustworthiness of Identification) (Q11)

The ultimate test of the effectiveness of automated threat classification is the degree to which users feel the need to double-check its output. Double-checking was used as an index of users' trust in the automation. Their trust was tapped in Question 11 through measures of the proportion of tracks that participants felt they needed to double-check for accuracy of classification.

The interpretation of the participants' ratings hinges on awareness that the scale is reversed, i.e., the most double-checking (lowest automation trust) is at the bottom of the axis and the least double-checking (highest automation trust) is at the top of the axis. Differences in users' judgments of the proportion of tracks that needed to be double-checked were not significant. Regardless of the kind of track (low threat or high threat) or declutter interface threshold (medium or high), participants reported a need to double-check some but not most of the tracks in all conditions (means were high threat/medium threshold = 2.93; high threat/high threshold = 2.94, low threat/medium threshold = 3.24, low threat/high threshold = 3.14). Analyses of the performance data should verify these perceptions of the users.

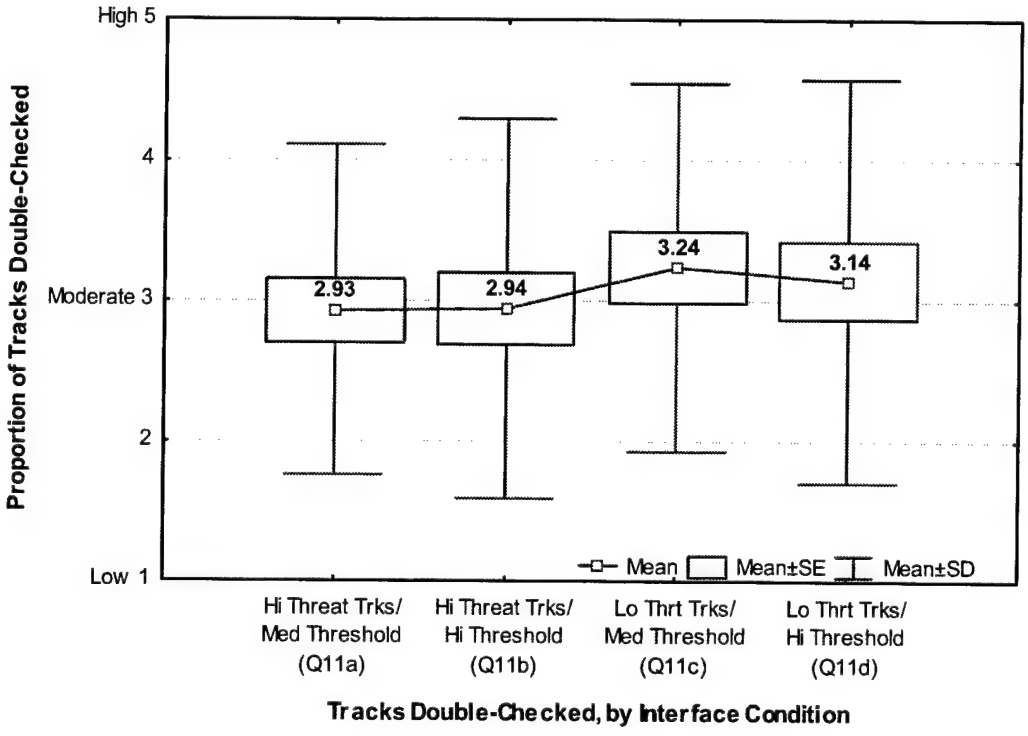

Figure 11. Automation trust as indexed by double-checking required.

STRATEGY FOR USING DECLUTTER (Q12).

Participants were asked to provide their strategy for using the declutter capabilities (Q12a) and whether or not they used different strategies for the medium and high threshold declutter interfaces (Q12b), with specification of how their strategies differed. Verbal interaction with the participants indicated that they were highly trained and socialized to be critical of automation. One participant said he believed in double-checking even himself, and then double-checking again. Most participants more or less expressed this attitude. When queried whether their strategies differed for high and medium threshold interfaces, 14 of the 27 participants reported that they used the same strategy regardless of the algorithm threshold; 10 participants reported they used different strategies depending on the declutter threshold; and 3 participants could not be classified, either because they did not answer or did not report whether or not their strategy differed.

Nature of Strategy for Using Declutter Capability (Q12a)

Participants reported a wide variety of strategies for using declutter capabilities. These strategies were commonly superimposed on basic strategies such as scanning from the center (close to ownship) outward. Many participants cited reliance on the declutter capability for prioritizing their attention to tracks, either in order (highlighted tracks attended to first) or frequency (checked decluttered tracks more / less often) or allocation of time ("checked highlighted tracks first, and more often, spent time on decluttered tracks"). Several cited the alerting value of declutter (change to highlighted when kinematics change). A small minority said they ignored decluttering ("didn't trust the program," "double-checking is a must").

Threshold-Related Differences in Strategies (Q12b)

Participants who differentiated their strategies, depending on the threshold setting for decluttering, reported a wide variety of strategies. Although still building on a close-in, then outward scan, several cited the need to "look harder at desaturated tracks (when using) the high threshold" interface ("higher threshold required more quality assurance (QA) of desaturated tracks") and the consequential heightened workload compared to the medium threshold interface. The general attitude seemed to be that "medium threshold declutter helped narrow down the tracks that were better candidates to recheck," while the "high threshold left me more suspicious of the decluttered tracks (causing) greater workload." Another person said "(he) acted on all high-threshold highlighted tracks and monitored some medium threat highlighted tracks." A minority of participants expressed distrust of the automation ("didn't trust high [threshold…because] it would dim a track as soon as it turned away and I would be more inclined to forget about that suspect track"). Some cited strategies of random checking of desaturated tracks, even though they placed higher priority on highlighted tracks.

RELATIVE VALUE OF DECLUTTER INTERFACES COMPARED TO NO DECLUTTER

Medium Threshold versus No Declutter Interface

Participants compared the declutter interfaces on six dimensions: procedural fit (Q13a), ease of learning (Q13c), ease of use (Q13e), efficiency of use (Q13g), effectiveness (Q13i), and quality of situation awareness (Q13k). They rated the medium threshold declutter interface substantially better than the no declutter interface on all six dimensions. The medium threshold declutter interface apparently fit well with their natural procedures (mean = 0.95), was relatively easy to learn (mean = 0.89) and especially easy to use (mean = 1.33), could be used efficiently (mean = 1.21) and effectively (mean = 1.26), and was seen as contributing substantially to the quality of their situation awareness (mean = 1.23).

High Threshold versus No Declutter Interface

Participants also compared the high threshold declutter interface with the no declutter interface on the same six dimensions (procedural fit (Q13b), ease of learning (Q13d), ease of use (Q13f), efficiency of use (Q13h), effectiveness (Q13j), and quality of situation awareness (Q13k)). Their results displayed a similar pattern with somewhat less positive ratings as their evaluations of the medium threshold declutter interface. The high threshold declutter interface was rated slightly better than the no declutter interface on all six dimensions, with means ranging from 0.63 to 0.87. Quality of situation awareness, ease of use, effectiveness, and quality of situation awareness received the highest ratings.

These results are consistent with the participants' rankings of the interfaces and their judgments of the overall usefulness of the interfaces, where the medium threshold declutter interface was rated superior to the high threshold declutter interface, and both interfaces were seen as superior to the no declutter interface.

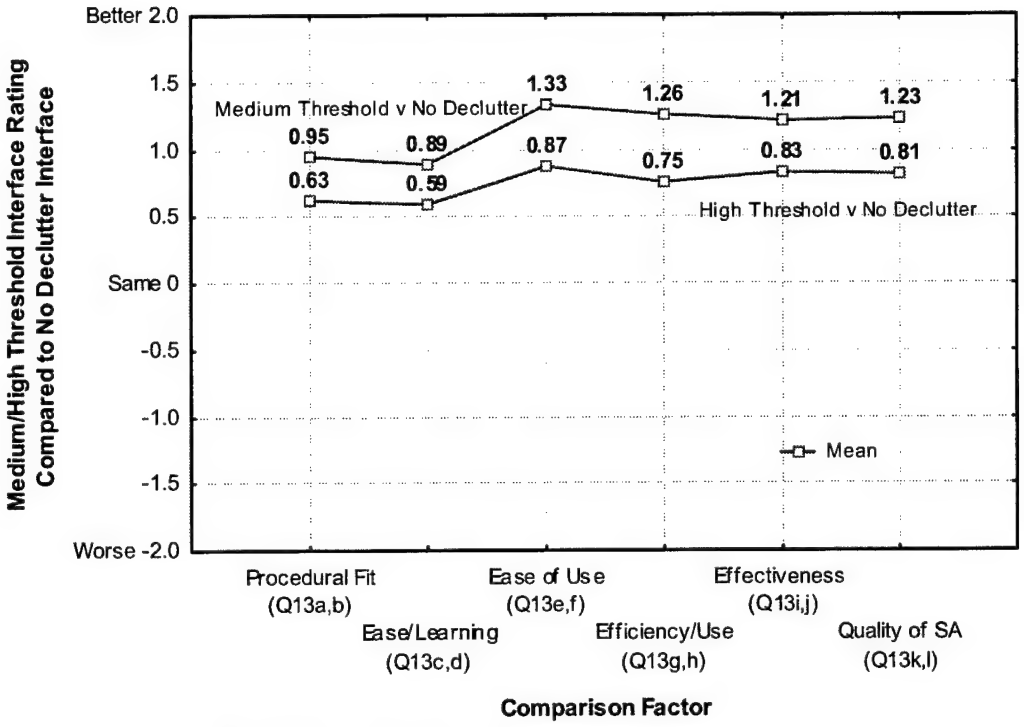

Figure 12. Declutter interfaces compared to no declutter interface.

OVERALL USEFULNESS RATINGS OF INTERFACES (Q14)

Participants' overall rating of the usefulness of the three interfaces placed the two declutter interfaces significantly higher than the no declutter interface ($p < 0.01$). The medium threshold declutter interface received the highest overall usefulness rating (mean = 4.13, marked by X_M in Figure 13). This interface was rated significantly more useful than either of the other two interfaces ($p = <0.05$). The high threshold declutter interface received the second highest overall usefulness rating (mean $[X_H] = 3.86$). The lowest usefulness was attributed to the no declutter interface (mean $[X_N] = 2.69$).

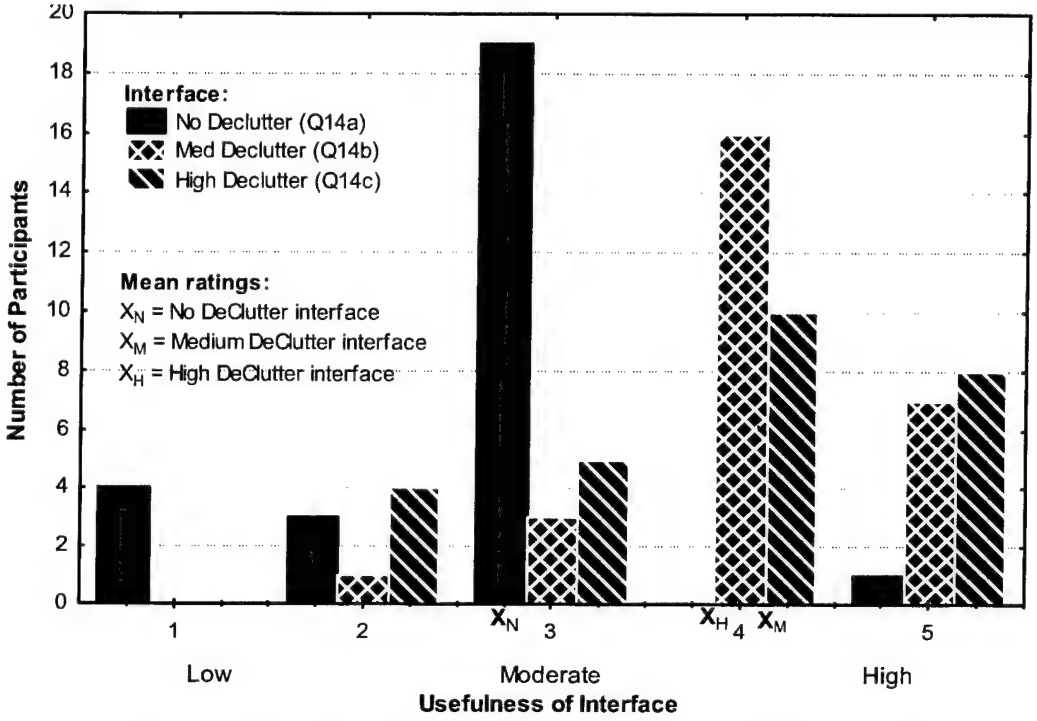

Figure 13. Distribution of overall usefulness ratings of interfaces.

INTERFACE PREFERENCES, RANK ORDER (Q15)

The participants' rank ordering of the three interfaces showed a clear preference for the medium declutter interface (19 of 27 participants), with only six participants preferring high declutter, and two preferring no declutter. Participants' second choice was predominantly high declutter, while their third choice was no declutter. Overall, 93% of the participants preferred some level of decluttering to no declutter, the equivalent of the traditional interface currently used in the Fleet.

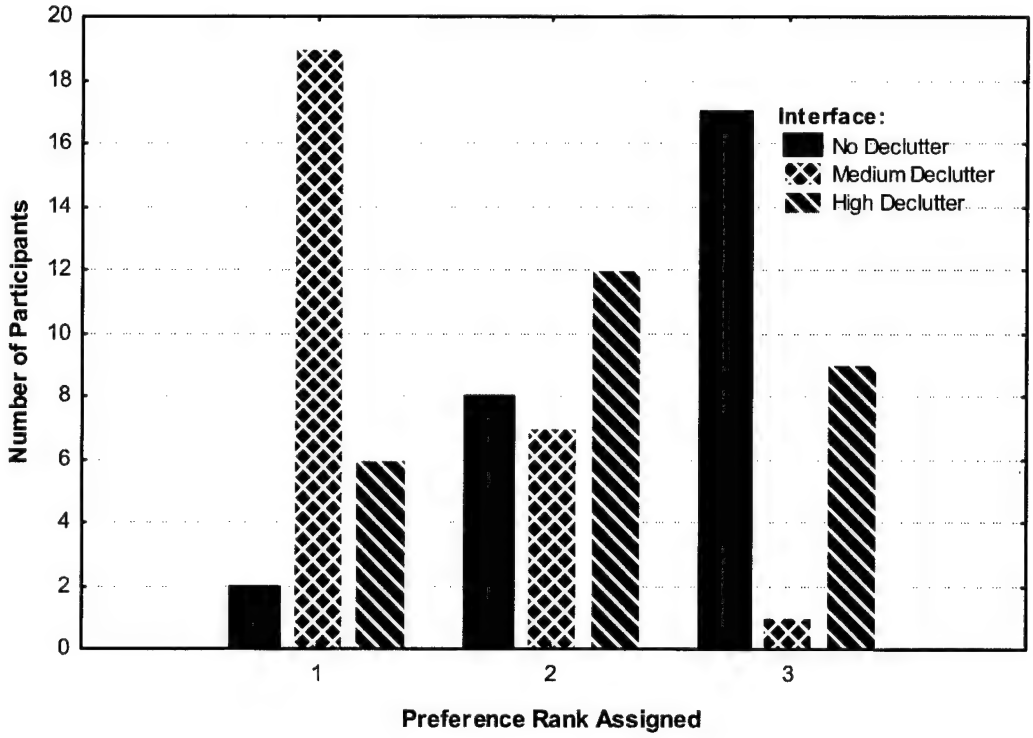

Figure 14. Preference ranking of interfaces.

DECLUTTER INTERFACE EVALUATIONS (Q17)

Strengths/Favorite Features

The most mentioned strength of the declutter interfaces was their ability to segregate contacts of less interest (e.g., ComAir) from those of more interest (e.g., threats, changing status), alerting users to the groupings, facilitating prioritizing, and contributing to quick evaluation and situation awareness. Participants liked having a "cleaner" picture with well-defined tracks while still being able to see all tracks, and they liked having ComAir lanes displayed so that they could instantly judge visually whether or not a track was travelling on an airlane. They liked having all track information available in a single block (our simulated character read out [CRO]), rather than having to navigate through complex paths to obtain single pieces of data about a track, as they must in their current Airborne Early Warning (AEW) operational system. They claimed that these features reduced their workload, relieved the pressure to act/decide quickly, allowed time to concentrate on suspects, and aided situation awareness. Several participants mentioned a preference for the medium threshold over the high threshold due to perceived benefits in workload, focusing attention and helping prioritize

tracks, and situation awareness. In essence, the reported workload benefits of the medium threshold were due to its encompassing relatively certain (high-score) threats, as well as "borderline" threats that a human decision-maker needed to check and monitor. The high threshold interface left the borderline threats undistinguished in the larger pool of non-threats, forcing users to double-check the entire non-threat pool to identify those threats judged to require continued monitoring. This situation was perceived as imposing a higher manual workload (more contacts to check), as well as a higher cognitive workload (mentally maintaining which of the decluttered tracks needed monitoring). This perceived additional workload was seen as reducing their overall efficiency. It was also seen as making it more difficult to maintain situation awareness as the situation evolved.

Weaknesses/Disliked Features

Participants mentioned various features they disliked, but there was little consensus. Their most frequent objection was the lack of change in the symbology when they acted on a track (notifying, querying, or warning), and the lack of user capability to add tags or labels (e.g., when they evaluated a track). Others objected to the lack of user-settable declutter criteria, incorporation of "real-world factors," and the need for a better understanding of the criteria used in the algorithm. Several objected to the performance of the algorithm on specific kinds of tracks. They also frequently cited concern over "loss of focus" on potential threat tracks and a "tendency to disregard decluttered tracks" or become "complacent," leading to avoidance of the necessity of checking all tracks. These concerns were expressed for themselves and for others. Others objected to the value of an algorithm that needed to be double-checked, generating a higher workload than necessary.

Numerous other features were cited only once or twice. These features included the need to train (users) to scan all tracks, basic track information unavailable by rollover and lack of territorial boundaries, special highlighting for changes in status, a track file log, "suspected hostile/bandit" symbology, and user capability to manually upgrade/degrade tracks.

Hard-to-Use Features

The general consensus was that the interfaces were easy to use. The only objections were the following: (1) having to use the mouse rather than key commands to take actions, which entailed rolling the mouse across the screen to the location of the action buttons, and (2) response tracking (keeping track of actions taken).

Suggestions for Tool/Interface Improvements

Participants offered many suggestions; nearly all were unique. Those features suggested by several respondents were as follows:

- Manually changeable symbols and/or colors for individual tracks
- Automatic change of symbol when track is acted on
- Selectable threshold for all criteria and weights in the declutter algorithm

Unique suggestions for additions were as follows:

- Settable range rings
- Procedures for sharing focus across the Battle Group and transitioning across sectors
- Tripwires for tasks that are due for decision by the user
- Extra factors in the algorithm to notify user of change in altitude or speed
- A user-assignable suspected hostile symbol
- User-selectable use/non-use of declutter so a team could have different views of GeoPlot

- Capability for the mission commander to designate tracks of interest (TOIs) that are highlighted on all displays
- Assignment of track numbers or tags to evaluated tracks
- Optimization of the threat classification algorithm—first by mutual agreement, then fine-tuning
- User to highlight own contacts of interest (COIs) by right-clicking
- Triggers for operator evaluation based on track parameters (e.g., rate of descent)
- Key commands
- Manual adjustment of overall declutter threshold
- Information provided by pre-hook (roll-over)
- Means of keeping track of contacts acted on

DISCUSSION

This experiment was designed to determine the value of GeoPlot decluttering in an operational setting. To accomplish this goal, we adopted a repeated measures design that yielded a comparative evaluation of the decluttered interfaces relative to a non-decluttered geoplot interface analogous to those currently common in the operational community. The core of the study was the users' performance with and without decluttering. We were also interested in understanding the workload implications of the interfaces, the automation trust issues, and the usability of the interface design. This report primarily addresses the usability question and touches on the topic of automation trust.

The design of this experiment, requiring each participant to use all three interfaces, allows them to make comparative judgments as well as absolute judgments of usability for performing the operational task. They judged the scenario and task to have a reasonable level of fidelity to actual operations. Based on their air defense warfare experience, they entered the study with moderately high confidence and expectations of low to moderate task difficulty. When initially exposed to the three different interfaces, they differentiated their confidence and expectations of task difficulty in an orderly fashion across the three interfaces, with greater confidence in the declutter interfaces and expectations that the declutter interfaces would make the task less difficult than the no declutter interface.

After experiencing the scenarios using the three interfaces, participants' confidence in their performance trended upward, especially for no declutter (which was initially lower than the declutter interfaces); however, participants reported the highest confidence in the declutter interfaces (especially medium threshold declutter) after experience using them to perform the air defense task. Their ratings of task difficulty using the no declutter interface were unaffected by task experience, but both declutter interfaces stabilized at a low to moderate level of rated task difficulty after the task.

Participants' subjective estimates of performance after experience in the scenarios were moderately high on three different measures: (1) threat detection, (2) change in threat status, and (3) ability to maintain situation awareness. In all cases, their subjective estimates of performance were higher for the declutter interfaces than for the no declutter interface. These estimates were supported by many verbal references to the beneficial impact of the declutter interfaces on workload.

When participants compared the declutter interfaces to the no declutter interface on six dimensions of usability, they gave the medium threshold declutter interface a greater advantage over no declutter than they gave the high threshold declutter interface. This response is consistent with their ratings of the usefulness of the interfaces, where both declutter interfaces were much more highly rated than the no declutter interface. It is also consistent with their preference rankings of the interfaces, where medium threshold declutter was strongly preferred to high threshold declutter, which was preferred to no declutter.

Participants were clearly aware of the issue of automation trust from a practical standpoint. They made many comments about agreement or disagreement with the algorithm, having to double-check its results, and its value in helping or hindering their attention to important tracks. We used two measures to tap their inclinations to trust or distrust the classification algorithm. The participants judged the accuracy of the algorithm in highlighting significant threats as quite high, possibly providing a necessary basis for trust. A second measure tapped participants' need to double-check the algorithm results by hooking tracks and making their own assessments. According to their reports on the need to double-check the automation, participants did not yield their complete trust to the automation, reporting that regardless of the declutter threshold or the automated classification, they

felt it necessary to check "some" of the tracks, as opposed to "few" or "most." These results, along with participants' comments about their training- and experience-based behavior patterns and their declutter usage strategies suggest that operationally experienced users appreciate the workload and prioritization benefits of automation while realistically tempering their reliance on it.

Participants claimed various creative strategies for using declutter capabilities. These strategies are probably subject to further refinement with extended experience, and they may take different form if additional declutter interface features that they requested are added. Although various rational strategies were used, most participants seemed to find the decluttering helpful in prioritizing their efforts, with further benefit (in most cases) in their perceived cognitive workload. Many additional features were suggested, and these suggestions and many verbal comments presume that the declutter interfaces are worthy of further development.

CONCLUSIONS

This study, which simulated an air defense task in a dense littoral operational environment, produced the following conclusions:

- Declutter interfaces were judged superior in overall usefulness compared to no declutter interface.
- Automation trust is realistic in this community, but the community is highly receptive to ways to prioritize work and reduce workload.
- Declutter interfaces rated higher usability than the no declutter interface on all measured dimensions.
- This evaluation of declutter interfaces compared with a no declutter interface established a clear mandate to build on the tested declutter concepts.
- Further development should incorporate additional features and tools suggested by these experienced air warfare personnel.

REFERENCES

Hutchins, S. G. 1996. "Principles for Intelligent Decision Aiding." SSC San Diego Technical Report 1718. Space and Naval Warfare Systems Center, San Diego, CA.

Hutchins, S. G., J. G. Morrison, and R. T. Kelly. 1996. "Principles for Aiding Complex Military Decision Making." *Proceedings of the Second International Symposium on Command and Control Research and Technology* (186–203). 25–28 June, Monterey, CA.

Kelly, R. T., S. G. Hutchins, and J. G. Morrison. 1996. "Decision Processes and Team Communications with a Decision Support System." *Proceedings of the Second International Symposium on Command and Control Research and Technology* (216–221). 25–28 June, Monterey, CA.

Liebhaber, M. J. 2001. "Description and Evaluation of an Air Defense Threat Assessment Algorithm." Technical Report 1887. Pacific Science and Engineering Group, Inc., San Diego, CA.

Liebhaber, M. J., D. A. Kobus, and B. A. Feher. 2002. "Studies of U.S. Navy Air Defense Threat Assessment: Cues, Information Order, and Impact of Conflicting Data." SSC San Diego Technical Report 1888. Space and Naval Warfare Systems Center, San Diego, CA.

Marshall, S. P., S. E. Christensen, and J. A. McAllister. 1996. "Cognitive Differences in Tactical Decision Making. *Proceedings of the 1996 Command and Control Research and Technology Symposium* (222–232). Monterey, CA.

Smallman, H. S., M. St. John, H. M. Oonk, and M. B. Cowen. 2001a. "Comparing Methods of Information Availability in 2-D and 3-D Displays," *IEEE: Computer Graphics and Application* (Special Issue on Applied Perception), *5*, 51–57.

Smallman, H. S., M. St. John, H. M. Oonk, and M. B. Cowen. 2001b. 'SYMBICONS': A Hybrid Symbology that Combines the Best Elements of SYMBols and ICONS." *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (110–114). Human Factors and Ergonomics Society, Santa Monica, CA.

St. John, M., Feher, B. A., and J. G. Morrison. 2002. "Evaluating Alternative Symbologies for Decluttering Geographical Displays." SSC San Diego TD 1890, Space and Naval Warfare Systems Center, San Diego, CA.

St. John, M., D. I. Manes, H. S. Smallman, B. A. Feher, and J. G. Morrison. 2004. "An Intelligent Threat Assessment Tool for Decluttering of Naval Air Defense Displays." SSC San Diego TR 1915, Space and Naval Warfare Systems Center, San Diego, CA.

APPENDIX A
USABILITY EVALUATION QUESTIONNAIRE

Usability Evaluation Questionnaire

Participant # _____ Date/Time: / /03

Post Introduction to Basic Task and Interface

1. Now that you are familiar with the general scenario – a fairly dense littoral environment - and your task (i.e. maintain situation awareness and respond to significant air threats in a timely manner), how **difficult** do you expect this task to be?

TASK DIFFICULTY	1----------2----------3----------4----------5 Low Moderate High

2. How **confident** are you that you can perform this task well?

CONFIDENCE	1----------2----------3----------4----------5 Low Moderate High

Post Introduction to All Three Interfaces

3. How **difficult** do you expect this task to be using each of the different interfaces?

a.	TASK DIFFICULTY Using the **No Declutter** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	TASK DIFFICULTY Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	TASK DIFFICULTY Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

4. How **confident** are you that you can perform this task well using each of the three interfaces?

a.	CONFIDENCE Using the **No Declutter** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	CONFIDENCE Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	CONFIDENCE Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

Debrief Following Test Scenarios

5. Now that you have completed the scenarios, how **difficult** was this task using each interface?

a.	TASK DIFFICULTY Using the **No Declutter** InterfaceLow	1----------2----------3----------4----------5 Moderate High
b.	TASK DIFFICULTY Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	TASK DIFFICULTY Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

6. How **confident** are you that you performed this task well using each interface?

a.	CONFIDENCE Using the **No Declutter** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	CONFIDENCE Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	CONFIDENCE Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

7. How would you rate your ability to **detect threats** while using each of the interfaces?

a.	THREAT DETECTION Using the **No Declutter** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	THREAT DETECTION Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	THREAT DETECTION Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

8. How would you rate your ability to **detect changes in threat status** while using each of the interfaces?

a.	DETECT CHANGES Using the **No Declutter** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	DETECT CHANGES Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	DETECT CHANGES Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

9. How would you rate your ability to **maintain overall situation awareness** while using each of the interfaces?

a.	MAINTAIN SITUATION AWARENESS Using the **No Declutter** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	MAINTAIN SITUATION AWARENESS Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
c.	MAINTAIN SITUATION AWARENESS Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

10. For the decluttered interfaces, how well did the automation succeed at "highlighting" all of the tracks that you rated to be significant threats (that is, never missing significant threats)?

a.	THREAT HIGHLIGHTING ACCURACY Using the **Medium Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High
b.	THREAT HIGHLIGHTING ACCURACY Using the **High Threshold** Interface	1----------2----------3----------4----------5 Low Moderate High

11. When using the Decluttered GeoPlot, what proportion of the tracks did you feel that you needed to **double-check** for accuracy of classification?

a.	HIGH THREAT TRACKS **Medium Threshold** Interface	1----------2----------3----------4----------5 Most Some Few
b.	HIGH THREAT TRACKS **High Threshold** Interface	1----------2----------3----------4----------5 Most Some Few
c.	LOW THREAT TRACKS **Medium Threshold** Interface	1----------2----------3----------4----------5 Most Some Few
d.	LOW THREAT TRACKS **High Threshold** Interface	1----------2----------3----------4----------5 Most Some Few

12. a. How did you use the declutter capabilities? What strategies did you use?_____

_____

_____

_____

_____

_____

_____

_____

12. b. Did your strategy for Medium and High Threshold differ? If so, how? _____

_____

_____

_____

_____

_____

_____

13. Taking the No Declutter Interface as the standard, what is your relative rating of the Decluttered Interfaces on the following dimensions?

	Medium Threshold Interface vs. No Declutter	High Threshold Interface vs. No Declutter
a-b. PROCEDURAL FIT (with your approach)	1----------2----------3----------4----------5 Worse Same Better	1----------2----------3----------4----------5 Worse Same Better
c-d. EASE OF LEARNING	1----------2----------3----------4----------5 Worse Same Better Worse	1----------2----------3----------4----------5 Same Better
e-f. EASE OF USE	1----------2----------3----------4----------5 Worse Same Better Worse	1----------2----------3----------4----------5 Same Better
g-h. TASK EFFICIENCY	1----------2----------3----------4----------5 Worse Same Better Worse	1----------2----------3----------4----------5 Same Better
i-j. EFFECTIVENESS	1----------2----------3----------4----------5 Worse Same Better Worse	1----------2----------3----------4----------5 Same Better
k-l. SITUATION AWARENESS	1----------2----------3----------4----------5 Worse Same Better	1----------2----------3----------4----------5 Worse Same Better

14. How would you rate the **overall usefulness** of the three interfaces in supporting your task?

a. OVERALL USEFULNESS1----------2----------3----------4----------5
Using the **No Declutter** InterfaceLow Moderate High

b. OVERALL USEFULNESS1----------2----------3----------4----------5
Using the **Medium Threshold** Interface Low Moderate High

c. OVERALL USEFULNESS1----------2----------3----------4----------5
Using the **High Threshold** Interface Low Moderate High

15. Which of the three interfaces would you **prefer to use** (Rank your preference—1=first choice; 3=third choice)?

___ **No Declutter** Interface
___ **Medium Threshold** Interface
___ **High Threshold** Interface

16. Rate the test **scenarios** on the following dimensions:

a. REALISM OF SCENARIO 1----------2----------3----------4----------5
Low Moderate High

b. REALISM OF TASK REQUIREMENTS 1----------2----------3----------4----------5
Low Moderate High

17. Evaluation of the **Decluttered Interfaces**:

a. Favorite Features / Strengths _____

b. Disliked Features / Weaknesses _____

c. Features That Were Hard To Use: _____

d. Suggestions For Tool / Interface Improvements: _____

APPENDIX B
ANALYSES BY AIR DEFENSE WARFARE EXPERIENCE

Air warfare expertise and experience of the 27 U.S. Navy participants was rated on a three-point scale for each participant by a subject matter expert. Fourteen of the participants were given a very high rating, two were given a high rating, and eleven were given a moderate rating. These categorizations were used in analyses to test for differences according to level of air warfare experience, where the Very High and High categories were combined to deal with the insufficient number (N = 2) of High participants; the combined Very High + High participants (N = 16) were contrasted with the Moderate participants (N = 11).

Although differences in participants' ratings according to their experience levels were generally not significant, participants who were classified as more experienced in air defense warfare assigned higher realism ratings to the scenarios than the less experienced participants ($mean_{Very\ High + High}$ participants = 3.73; $mean_{Moderate}$ participants = 3.14; p = 0.02). More experienced participants also tended to rate the experimental task higher in realism than less experienced participants ($mean_{VH+H}$ = 3.86; $mean_{M}$ = 3.27; p = 05).

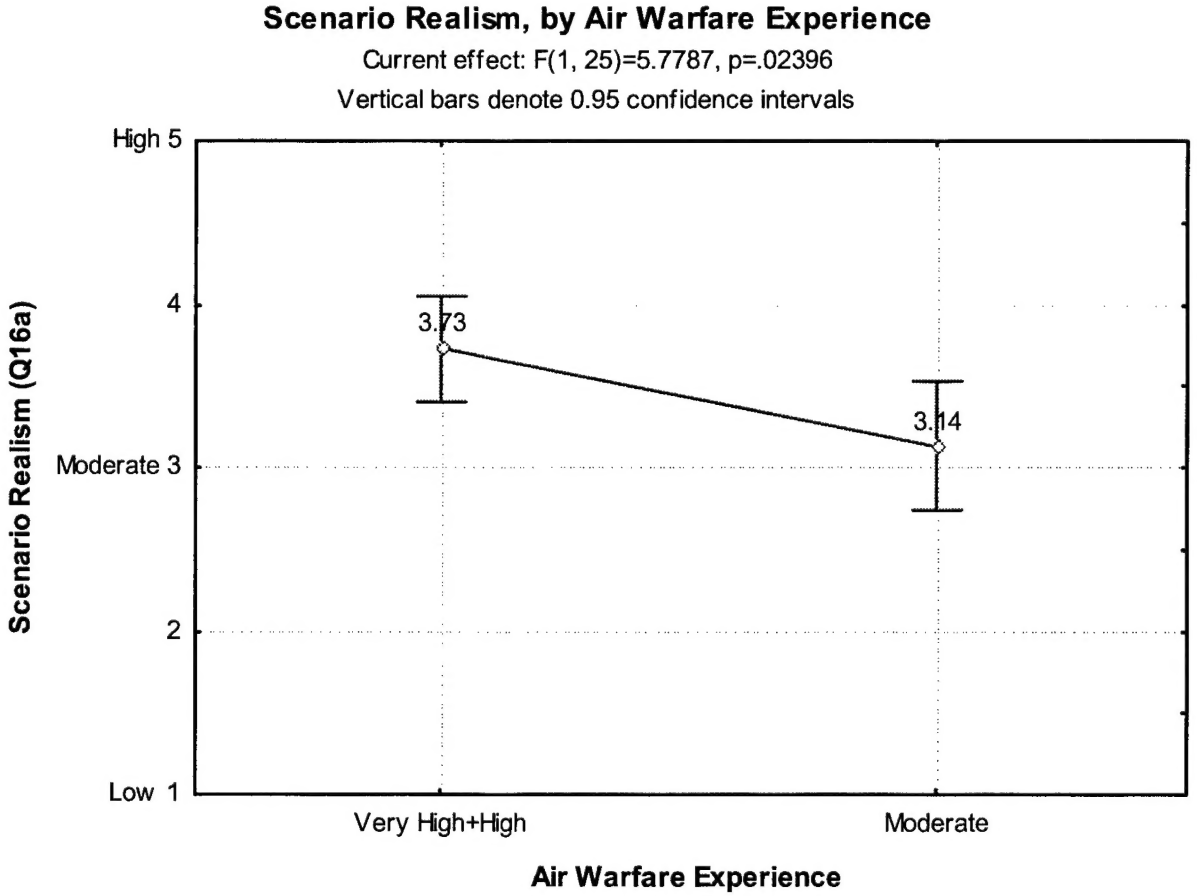

Figure B-1. Scenario realism ratings by participants categorized by their air warfare experience.

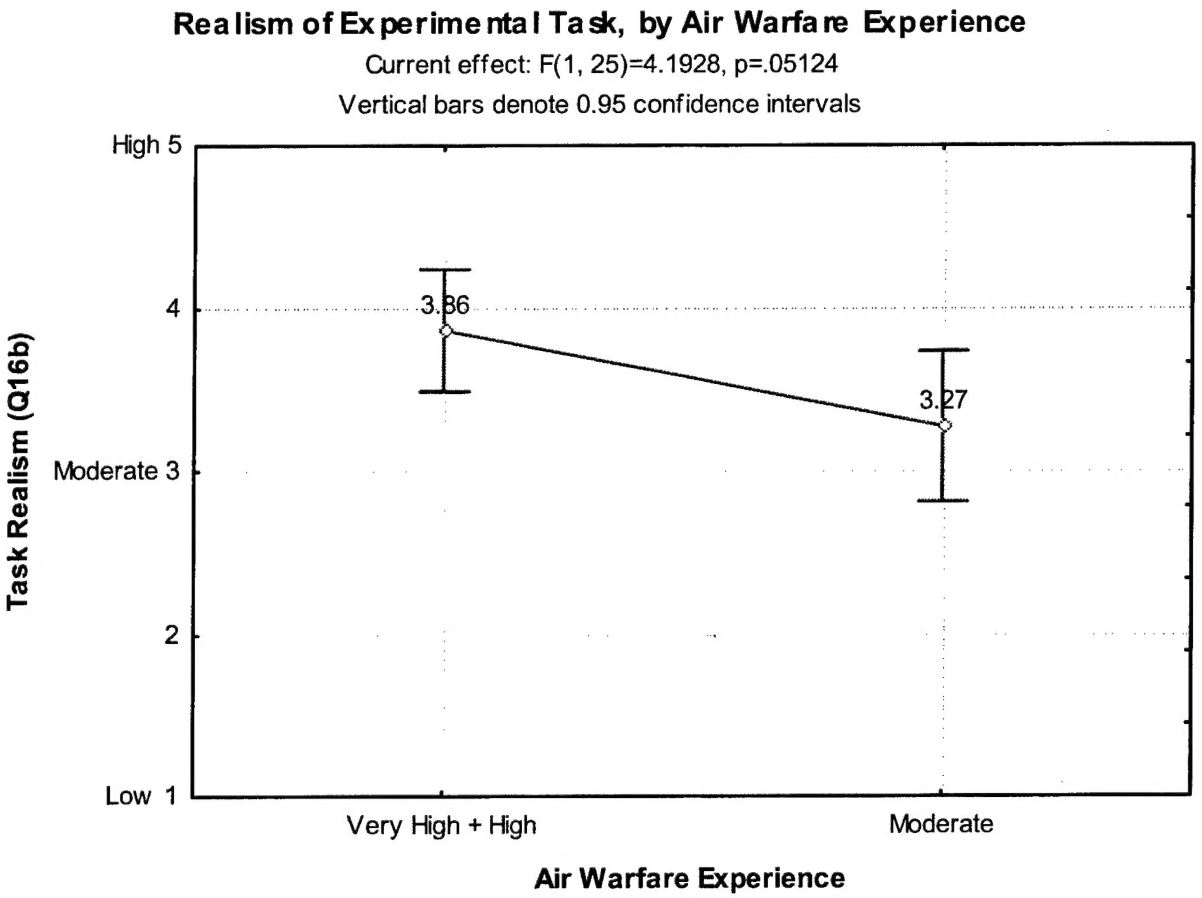

Figure B-2. Experimental task realism ratings by participants categorized by their air warfare experience.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-01-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
07–2004	Technical	

4. TITLE AND SUBTITLE

GEOPLOT DECLUTTER INTERFACE USABILITY EVALUATION

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0602233N

6. AUTHORS

Bela Feher

5d. PROJECT NUMBER

5e. TASK NUMBER
03328

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SSC San Diego
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER
TR 1917

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research, Code 342
800 N. Quincy St.
Arlington, VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This study evaluates the usability of an advanced human–system interface that uses an automated algorithm to reduce user workload by automatically decluttering low-threat tracks. The usability evaluation was conducted in association with an experiment conducted for two purposes: (1) to compare two decluttered interfaces with a standard, non-decluttered interface for performing an air defense task, and (2) to compare medium and high threshold cutoffs on the declutter algorithm.

15. SUBJECT TERMS

Mission Area: Human Factors Engineering

naval air defense	threat assessment	declutter algorithm
situation awareness	threat categories	usability

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Bela Feher
U	U	U	UU	52	19B. TELEPHONE NUMBER *(Include area code)* (619) 553–9226

INITIAL DISTRIBUTION

20012	Patent Counsel	(1)
21513	Archive/Stock	(4)
21512	Library	(2)
215	G. C. Pennoyer	(1)
2151	F. F. Roessler	(1)
21513	D. Richter	(1)
240	R. Jaffee	(1)
244	J. L. Martin	(1)
2441	J. G. Morrison	(7)
244210	B. Feher	(8)
244210	R. Smillie	(1)

Defense Technical Information Center
Fort Belvoir, VA 22060-6218 (4)

SSC San Diego Liaison Office
C/O PEO-SCS
Arlington, VA 22202-4804 (1)

Center for Naval Analyses
Alexandria, VA 22311-1850 (1)

Office of Naval Research
ATTN: NARDIC
Philadelphia, PA 19111-5078 (1)

Government-Industry Data Exchange
Program Operations Center
Corona, CA 91718-8000 (1)

Pacific Science & Engineering Group
San Diego, CA 92121 (3)